International Journal of Mathematics and Statistics

ISSN (Online): 2208-2212
Volume 09 Issue 04 December 2023

DOI: https://doi.org/10.69980/917v0q46

TEMPORAL-RESOLUTION EFFECTS ON MULTIVARIATE CALIBRATION AND AGREEMENT OF LOW-COST AIR QUALITY SENSORS

Dr. Thomas R. Ellington[1*], Prof. Mei-Lin K. Zhang[2], Dr. Javier M. Ortega[3]

[1*]*Department of Environmental Statistics and Air Quality Science, Westbridge Institute of Environmental Technology, London, United Kingdom*
[2]*School of Atmospheric Measurement and Data Analytics, Pacific Institute of Environmental Research, Vancouver, Canada*
[3]*Department of Applied Statistics and Environmental Monitoring, Iberian University of Science and Technology, Valencia, Spain*

Abstract:

Extension of monitoring coverage may be achieved using low-cost air quality sensors, but this may lead to systematic differences in their measurements from reference sensors and/or affect their variability, calibration accuracy, and measurement agreement; temporal aggregation may affect their variability, calibration accuracy, and measurement agreement. This study assessed how 10-, 30-, and 60-minute temporal resolutions influence multivariate calibration, cross-sensor agreement, and chronological predictive performance for NO_2 and O_3 measurements. Correspondence of paired reference and low-cost sensor observations was assessed by the time stamp, descriptive statistics, Pearson and Spearman correlations, simple and multivariate ordinary least-squares calibration, Bland–Altman agreement, and proportional-bias analysis. Environmental covariates (temperature and relative humidity) were included, and heteroscedasticity and autocorrelation were addressed by using Newey–West inference. Temporal generalization was evaluated using chronological 80:20 validation. Weak negative cross-sensor correlations were found for NO_2, while the correlation of O_3 was found to be increasing with time from 0.352 after 10 minutes to 0.446 after 60 minutes. A greater improvement was obtained for O_3 using multivariate calibration, and its highest R^2 value was obtained at 30 minutes, with an R^2 value of 0.382. Temporal aggregation generally decreased calibration errors and narrowed the agreement limits, although O_3 retained substantial systematic bias. Chronological test R^2 remained near zero for NO_2 and negative for O_3 across all temporal resolutions. These results show that temporal aggregation can improve statistical stability and calibration accuracy without necessarily improving temporal generalisation.

Keywords: *low-cost sensors, temporal aggregation, multivariate calibration, Bland–Altman agreement, air-quality monitoring*

1. Introduction:

Air-quality monitoring is essential to understand the nature of air pollution, to track changes in pollutant levels with time, and to inform environmental management decisions based on science. Regulatory monitoring stations are expensive to install, operate and maintain and offer high-quality measurements, but can limit the number of monitoring stations. In recent years, the Internet of Things (IoT) has emerged to provide opportunities for spatially distributed and continuous monitoring, where modern architectures combine sensor networks, data transmission, data processing and analytics, and automations for air-quality assessment systems (Garcia et al., 2025) Low-cost sensors (LCSs) have thus gained significant importance as an additional monitoring technology, as they potentially enable the observation of the atmosphere over longer periods and with a much larger deployment density than conventional instrumentation (Liu et al., 2020). Equivalence with reference measurements does not necessarily imply greater accessibility. Low-cost air-quality sensors have been compared in terms of their performance with varying degrees of success, showing significant differences in accuracy between sensor types, pollutants, operating conditions and evaluation settings (Karagulian et al., 2019). This variability is especially relevant when the concentration changes are represented by the measurements made by the sensors. The concentrations of atmospheric pollutants are the result of dynamic interaction between emissions, dispersion, meteorological and local environmental processes. Mathematical representation of industrial emissions is shown to be important in quantifying these processes when evaluating the quality of the atmosphere, because the dependence of the atmospheric-air quality on the industrial emissions is clearly illustrated by mathematical modeling of these emissions (Lukyanets et al., 2023). The role of observation- and computation-based frameworks of increasing detail for describing changes in atmospheric composition, both temporally and spatially, has also been highlighted in the advancement of air-quality modelling and forecasting (Baklanov & Zhang, 2020). A reliable use of the LCS observations will thus necessitate a statistical evaluation of the closeness of the observations to reference observations for different situations.

The focus of this assessment is calibration, since the raw sensor response may be influenced by the environment and/or the measurement properties of the sensor. Calibration, environmental sensitivity, sensor aging, deployment conditions, and data quality procedures are still key factors in the performance of measurements with outdoor LCS deployments (Okorn & Iraci, 2024). Multivariate statistical modeling offers a useful method that uses pollutant concentrations as part of a model that also includes applicable environmental conditions as predictors. This research approach is broadly supported by spatiotemporal air-quality research, where multivariate regression has been used to quantify relationships between air-quality pollutant concentrations and climatic variables, and by seasonal variations (Tella et al., 2021). The use of more complex machine-learning methods for modelling outdoor air quality is also becoming more common, but the benefits of such models should be evaluated against the interpretability of the model and the nature of the measurements used to model the quality (Rybarczyk & Zalakeviciute, 2018).

Assessing the performance of the calibration also means distinguishing between association and agreement. A high correlation does not necessarily mean that the two series of measurements are similar enough. To address this difference Bland–Altman analysis quantifies mean differences and limits of agreement and can be systemically and quantitatively examined in terms of mean differences and limits of agreement between the measurement methods (Doğan, 2018). The distinction is important for LCS applications since the selection, deployment strategy and data quality features of the measurements could significantly affect the suitability of sensor observations for the following exposure modeling (Bi et al., 2022). Previous research into low-cost monitoring that could be applied in practice also showed that calibration, validation, data interpretation, and characterization of sensor performance were all key requirements for the use of sensors in practice (Clements et al., 2017).

A statistical view on calibration and agreement can be added thanks to temporal resolution. The longer the time scale of the aggregation the less likely it is that short term variability or measurement noise will influence the result, but the longer the time scale the greater the chance that the correlation structure, calibration error, and apparent agreement with the reference instruments will change. The widespread adoption of low-cost sensors for exposure assessment and air quality led to an understanding that characteristics of measurements impact interpretation of the information provided by the sensor (Morawska et al., 2018). This is especially true for gaseous pollutants like NO_2 and O_3. Low-cost NO_2 and O_3 stations have been calibrated and validated and have proven useful for monitoring networks if their responses are properly characterized and calibrated to reference readings (Cavaliere et al., 2023).

Although significant advances have been made in the assessment of LCS, the accuracy of calibration, cross-sensor association, agreement of measurements and temporal resolution are still treated separately. There has been less focus on the consistency of changes resulting from temporal aggregation across the three measures (correlation, multivariate calibration, agreement, and out-of-sample). When temporally ordered environmental observations are heteroscedastic and serially dependent, special caution must also be exercised in making statistical inferences. Thus, a combination of statistical regression inference and agreement analysis and chronological validation can give a more comprehensive evaluation than the use of correlation or in-sample goodness of fit.

The objective of this study is to quantify the dependence and agreement between low-cost and reference NO_2 and O_3 measurements at various temporal resolutions; to assess the performance of robust multivariate calibration including pollutant measurements, temperature and relative humidity; and to quantify the effect of temporal aggregation on calibration accuracy, agreement between the low-cost and reference measurements, and chronological predictive performance at various timescales (10, 30, 60 minutes).

2. Materials and Methods

2.1 Research Design

The study used a quantitative statistical calibration design to assess agreement between measurements of air quality from low-cost air quality sensors and the reference monitoring observations. The analysis was focused on the NO_2 and O_3 pollutants as both pollutants were available in paired form for the temporal resolutions of 10-minute, 30-minute and 60-minute. The methodology involved testing distributional behaviour, cross-sensor association, multivariate calibration, model stability, agreement and the impact of temporal aggregation. Focus was given to statistical methods of statistics that are easily understood and not overly complicated, such as black-box prediction.

2.2 Data Source

All the data were taken from paired measurements of regulatory air-quality monitoring and low-cost air-quality monitoring in Bulevar Sur, Valencia, Spain (Meneses Albalá et al., 2025). The files included measurements of pollutants and meteorological data, both with resolutions of 10, 30, and 60 minutes, and they were synchronized. The variables used in the study were the reference and low-cost measurements of NO_2 and O_3, temperature and relative humidity.

2.3 Data Preparation and Validity Screening

All files were matched based on the recorded timestamps and not the row position. Timestamps within each source were duplicated and combined by their arithmetic mean for each time point and paired. Measurements that were not available at all in the reference measurements were not included. Negative concentrations of NO_2 were excluded as invalid and filled in with missing data. Observations of NO_2 with negative values were not considered valid and were kept as missing data. Then analyses were performed on the available paired observations without any imputations of missing reference pollutant values.

2.4 Descriptive and Correlation Analysis

Descriptive statistics were computed for each temporal resolution for both the reference and low-cost NO_2 and O_3 sensors. These included the number of samples, mean, median, standard deviation, minimum, quartiles, maximum, coefficient of variation, skewness and kurtosis. Two types of association were used, Pearson correlation and Spearman rank correlation, for cross-sensor association. Pearson coefficients were used to assess linear relationships, while the Spearman coefficients were used to assess a rank-based measure which is less affected by distributional asymmetry. Changes due to aggregation were quantified by making correlations between different temporal resolutions.

2.5 Statistical Calibration Models

For each pollutant, the two pollutant calibration specifications were examined for each temporal resolution. The simple model only relied on the low-cost pollutant measurement while the multivariate model also included the low-cost sensor temperature and relative humidity. Ordinary least-squares regression was used to estimate model coefficients. The R2, adjusted R2, root mean square error and mean absolute error were used to summarize the model performance. To assess if the environmental covariates were able to explain anything beyond the simple calibration model, nested F-tests were performed.

2.6 Model Diagnostics and Robust Inference

Variance inflation factors, Breusch-Pagan test, residual-versus-fitted plots, Q-Q plots and the Durbin-Watson statistic were used to analyze regression assumptions. The predictor collinearity was checked using variance inflation factors and the heteroscedasticity was assessed using Breusch-Pagan statistics and the serial dependence using Durbin-Watson statistics. Coefficient inferences were based on Newey-West standard errors that are heteroscedasticity- and autocorrelation-consistent, due to the presence of non-constant variance and high levels of autocorrelation in the residuals. Hence, strong confidence intervals and significance tests were provided for the final multivariate models.

2.7 Agreement and Proportional-Bias Analysis

Bland-Altman statistics were used to determine the agreement between the low cost and reference measurement. The mean difference between low-cost and reference measurements, the standard deviation of the difference and the limits of agreement (LoA95%) for each pollutant and temporal resolution were determined. Also considered was mean absolute difference. The difference between the sensor value and the reference value was regressed against the mean value of the two values to detect the presence of proportional bias. If there was a non-zero slope the size of the disagreement was systematic over the range of measurement.

2.8 Chronological Validation and Temporal-Resolution Assessment

Model generalization was evaluated with an 80:20 split of the data set, where the first 80% of data observations were used to model the system, and the last 20% were used to test the model. It was important that the information was not leaked by the temporal order, so this was preserved. The R^2, RMSE and MAE were used to assess test performance. Finally, the results for 10, 30 and 60 minutes were compared to assess the effect of temporal aggregation on the statistical calibration of variability, correlation, calibration error, agreement width, proportional bias and out-of-sample performance.

3. Results

3.1 Characteristics of the Student Sample

Table 1 shows the descriptive characteristics of NO_2 and O_3 measured by the reference station and low-cost sensor (LCS). Reference NO_2 was similar at 10-, 30-, and 60-minute resolutions from 14.463 to 14.508; and LCS NO_2 was near 15.24 across all resolutions. All LCS O_3 means were significantly greater than the reference O_3 means. Standard deviations generally decreased with temporal aggregation, particularly for LCS O_3. Reference NO_2 was positively skewed across all temporal resolutions, whereas LCS NO_2 showed consistently negative skewness.

Table 1. Descriptive statistics across temporal resolutions

Resolution (min)	Variable	N	Mean	Median	SD	Minimum	Q1	Q3	Maximum	CV (%)	Skewness	Kurtosis
10	LCS NO_2	22,979	15.250	18.270	6.016	0.747	15.045	18.816	18.816	39.450	-1.563	0.836
30	LCS NO_2	7,666	15.243	18.264	5.852	0.764	14.423	18.816	18.816	38.390	-1.579	0.982
60	LCS NO_2	3,838	15.239	18.223	5.701	0.787	14.240	18.816	18.816	37.409	-1.602	1.145
10	LCS O_3	22,979	112.377	90.444	73.610	39.246	46.738	158.263	284.787	65.503	0.893	-0.300
30	LCS O_3	7,666	113.013	96.374	68.538	39.246	50.135	161.463	261.633	60.646	0.709	-0.675
60	LCS O_3	3,838	113.307	97.842	66.504	39.579	51.728	163.254	255.764	58.694	0.648	-0.764
10	Reference NO_2	21,490	14.463	9.000	14.702	0.000	4.000	19.000	97.000	101.656	1.934	3.983
30	Reference NO_2	7,161	14.476	9.000	14.339	0.000	4.000	18.667	92.333	99.056	1.922	3.954
60	Reference NO_2	3,581	14.508	9.143	14.047	1.200	4.200	18.800	92.800	96.819	1.908	3.900
10	Reference O_3	22,979	52.418	58.000	26.027	3.000	32.000	72.000	92.000	49.653	-0.439	-0.903
30	Reference O_3	7,666	54.881	59.667	24.493	8.000	37.000	72.667	96.000	44.630	-0.379	-0.780
60	Reference O_3	3,838	55.718	60.400	25.013	8.714	37.650	73.400	97.854	44.892	-0.352	-0.779

3.2 Cross-Sensor Correlation

Table 2 shows weak negative correlations between reference and LCS NO_2 at all resolutions, with Pearson coefficients between −0.105 and −0.115. O_3 showed stronger positive correlations, increasing from 0.352 at 10 minutes to 0.446 at 60 minutes. Spearman coefficients followed a similar pattern. All correlations were statistically significant at $p < 0.001$. Figure 1 shows the change in cross-sensor association with temporal aggregation. Figure 1A shows Pearson correlations for NO_2 and O_3 across the three resolutions. Figure 1B shows the corresponding Spearman rank correlations.

Table 2. Cross-sensor correlations

Resolution (min)	Pollutant	N	Pearson r	Pearson p	Spearman ρ	Spearman p
10	NO_2	21,490	-0.105	<0.001	-0.111	<0.001
30	NO_2	7,161	-0.112	<0.001	-0.131	<0.001
60	NO_2	3,581	-0.115	<0.001	-0.152	<0.001
10	O_3	22,979	0.352	<0.001	0.360	<0.001
30	O_3	7,666	0.443	<0.001	0.441	<0.001
60	O_3	3,838	0.446	<0.001	0.451	<0.001

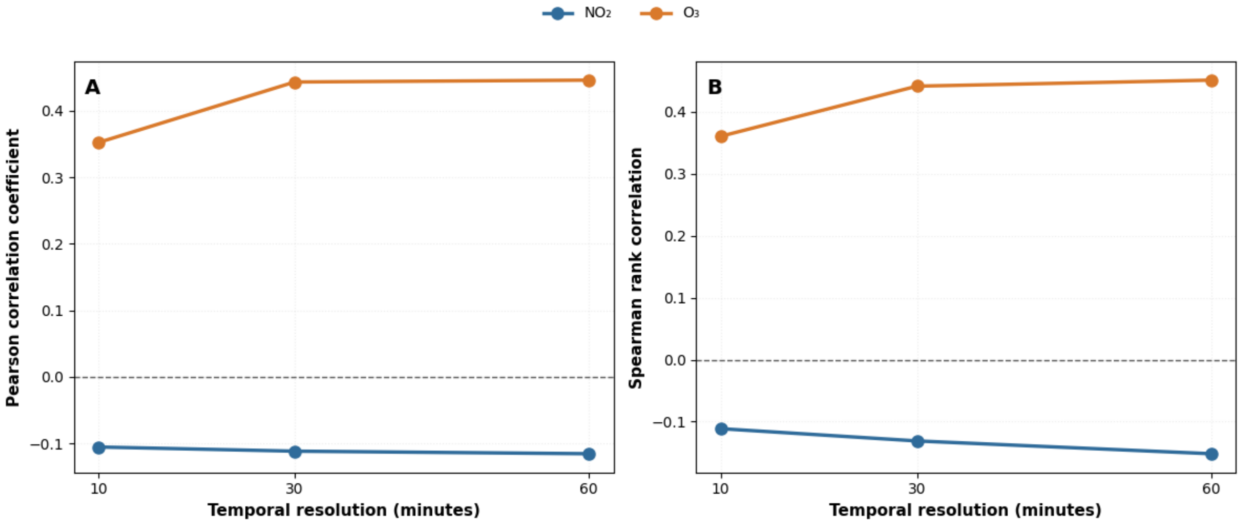

Figure 1. Cross-Sensor Correlations (A) Pearson correlation; (B) Spearman correlation.

3.3 Calibration Model Performance

Table 3 shows that multivariate calibration improved model performance compared with simple calibration for both pollutants. For NO_2, multivariate R^2 increased from 0.056 at 10 minutes to 0.065 at 60 minutes, while RMSE declined from 14.280 to 13.579. O_3 showed larger improvements, with the highest multivariate R^2 of 0.382 at 30 minutes. At the same resolution, RMSE decreased from 21.952 to 19.258 and MAE from 18.362 to 15.615. All nested model comparisons were significant at $p < 0.001$. Figure 2 shows the error differences between simple and multivariate calibration models. Figure 2A shows NO_2 RMSE across temporal resolutions. Figure 2B shows NO_2 MAE across temporal resolutions. Figure 2C shows O_3 RMSE across temporal resolutions. Figure 2D shows O_3 MAE across temporal resolutions.

Table 3. Simple and multivariate calibration performance

Resolution (min)	Pollutant	N	Simple R^2	Multivariate R^2	ΔR^2	Simple RMSE	Multivariate RMSE	RMSE improvement (%)	Simple MAE	Multivariate MAE	MAE improvement (%)	Nested F	Nested p
10	NO_2	21,490	0.011	0.056	0.045	14.620	14.280	2.323	10.732	10.410	3.002	516.951	<0.001
30	NO_2	7,161	0.012	0.061	0.048	14.248	13.894	2.485	10.458	10.129	3.152	184.681	<0.001
60	NO_2	3,581	0.013	0.065	0.052	13.951	13.579	2.666	10.223	9.881	3.342	99.312	<0.001
10	O_3	22,979	0.124	0.292	0.168	24.356	21.902	10.078	20.356	17.448	14.284	2719.086	<0.001
30	O_3	7,666	0.197	0.382	0.185	21.952	19.258	12.274	18.362	15.615	14.959	1146.982	<0.001
60	O_3	3,838	0.199	0.370	0.171	22.383	19.848	11.323	18.651	15.906	14.717	520.806	<0.001

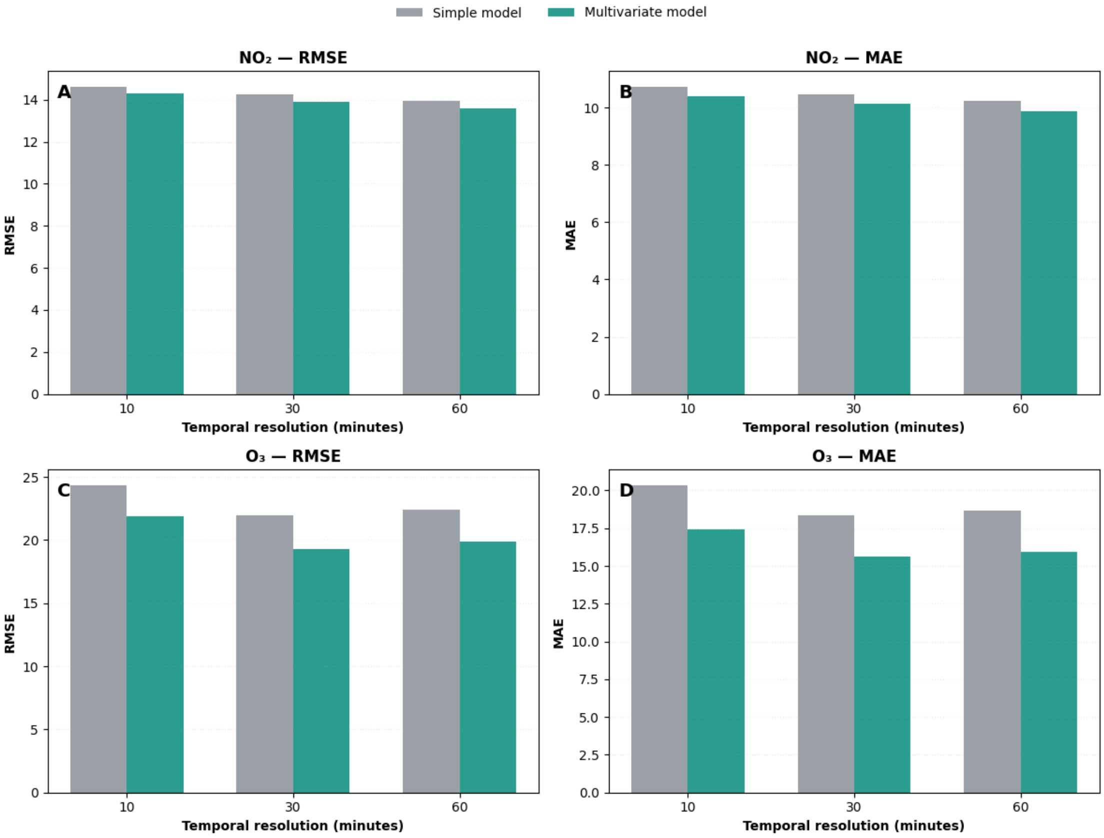

Figure 2. Calibration Error Comparison: (A) NO_2 RMSE; (B) NO_2 MAE; (C) O_3 RMSE; (D) O_3 MAE.

3.4 HAC-Robust Calibration Coefficients

Table 4 shows the HAC-robust multivariate calibration coefficients. LCS NO_2 coefficients were negative at all resolutions, ranging from −0.238 to −0.266. Temperature also showed negative coefficients, while humidity was positively associated with reference NO_2. For O_3, LCS coefficients were positive and increased from 0.076 at 10 minutes to 0.105 at 60 minutes. Temperature coefficients were positive, whereas humidity coefficients remained negative. The major predictor effects remained statistically significant after HAC adjustment.

Table 4. HAC-robust calibration coefficients

Resolution (min)	Pollutant	N	HAC lag	Term	Coefficient	HAC SE	t	p	CI lower	CI upper
10	NO_2	21,490	13	Humidity	0.114	0.022	5.136	<0.001	0.071	0.158
10	NO_2	21,490	13	Intercept	12.708	3.222	3.944	<0.001	6.393	19.024
10	NO_2	21,490	13	LCS NO_2	-0.238	0.047	-5.028	<0.001	-0.331	-0.145
10	NO_2	21,490	13	Temperature	-0.245	0.073	-3.375	<0.001	-0.387	-0.103
30	NO_2	7,161	10	Humidity	0.116	0.029	4.081	<0.001	0.061	0.172
30	NO_2	7,161	10	Intercept	12.752	4.175	3.054	0.002	4.567	20.937
30	NO_2	7,161	10	LCS NO_2	-0.254	0.062	-4.089	<0.001	-0.376	-0.132
30	NO_2	7,161	10	Temperature	-0.244	0.096	-2.552	0.011	-0.432	-0.057
60	NO_2	3,581	8	Humidity	0.119	0.031	3.802	<0.001	0.058	0.181
60	NO_2	3,581	8	Intercept	12.793	4.639	2.758	0.006	3.697	21.889
60	NO_2	3,581	8	LCS NO_2	-0.266	0.072	-3.688	<0.001	-0.407	-0.124
60	NO_2	3,581	8	Temperature	-0.248	0.108	-2.298	0.022	-0.460	-0.036
10	O_3	22,979	13	Humidity	-0.496	0.032	-15.364	<0.001	-0.559	-0.433
10	O_3	22,979	13	Intercept	77.890	4.583	16.996	<0.001	68.907	86.872

10	O_3	22,979	13	LCS O_3	0.076	0.007	11.421	<0.001	0.063	0.089
10	O_3	22,979	13	Temperature	0.544	0.116	4.701	<0.001	0.317	0.771
30	O_3	7,666	10	Humidity	-0.486	0.041	-11.818	<0.001	-0.567	-0.405
30	O_3	7,666	10	Intercept	75.530	5.929	12.739	<0.001	63.907	87.153
30	O_3	7,666	10	LCS O_3	0.098	0.009	11.254	<0.001	0.081	0.115
30	O_3	7,666	10	Temperature	0.610	0.140	4.357	<0.001	0.335	0.884
60	O_3	3,838	8	Humidity	-0.487	0.050	-9.746	<0.001	-0.585	-0.389
60	O_3	3,838	8	Intercept	76.137	7.256	10.493	<0.001	61.910	90.363
60	O_3	3,838	8	LCS O_3	0.105	0.012	9.018	<0.001	0.082	0.128
60	O_3	3,838	8	Temperature	0.586	0.180	3.266	0.001	0.234	0.938

3.5 Agreement and Temporal-Resolution Effects

Table 5 shows the combined calibration, agreement, and validation results. NO_2 mean bias declined from 0.745 at 10 minutes to 0.685 at 60 minutes, while calibration RMSE decreased from 14.280 to 13.579. O_3 mean bias remained substantially larger, ranging from 59.958 to 57.589. O_3 calibration R^2 was highest at 30 minutes (0.382). Test R^2 values were low for NO_2 and negative for O_3, while test RMSE declined as temporal aggregation increased. Figure 3 shows the agreement patterns across temporal resolutions. Figure 3A shows NO_2 agreement at 10 minutes. Figure 3B shows O_3 agreement at 10 minutes. Figure 3C shows NO_2 agreement at 30 minutes. Figure 3D shows O_3 agreement at 30 minutes. Figure 3E shows NO_2 agreement at 60 minutes. Figure 3F shows O_3 agreement at 60 minutes.

Table 5. Integrated calibration and agreement results

Pollutant	Resolution (min)	Reference N	Reference SD	Reference CV (%)	Pearson r	Spearman ρ	Calibration R^2	Calibration RMSE	Calibration MAE	Mean bias	Lower 95% LoA	Upper 95% LoA	Proportional-bias slope	Bias-model R^2	Test R^2	Test RMSE	Test MAE
NO_2	10	21,490	14.702	101.656	-0.105	-0.111	0.056	14.280	10.410	0.745	-31.560	33.051	-1.533	0.506	0.010	18.705	14.231
NO_2	30	7,161	14.339	99.056	-0.112	-0.131	0.061	13.894	10.129	0.723	-30.838	32.284	-1.542	0.508	0.012	18.411	14.053
NO_2	60	3,581	14.047	96.819	-0.115	-0.152	0.065	13.579	9.881	0.685	-30.242	31.612	-1.552	0.512	0.015	18.247	13.934
O_3	10	22,979	26.027	49.653	0.352	0.360	0.292	21.902	17.448	59.958	-75.059	194.975	1.273	0.636	-0.106	24.188	20.692
O_3	30	7,666	24.493	44.630	0.443	0.441	0.382	19.258	15.615	58.133	-62.829	179.095	1.208	0.650	-0.155	23.096	19.839
O_3	60	3,838	25.013	44.892	0.446	0.451	0.370	19.848	15.906	57.589	-59.423	174.601	1.162	0.619	-0.124	22.369	19.134

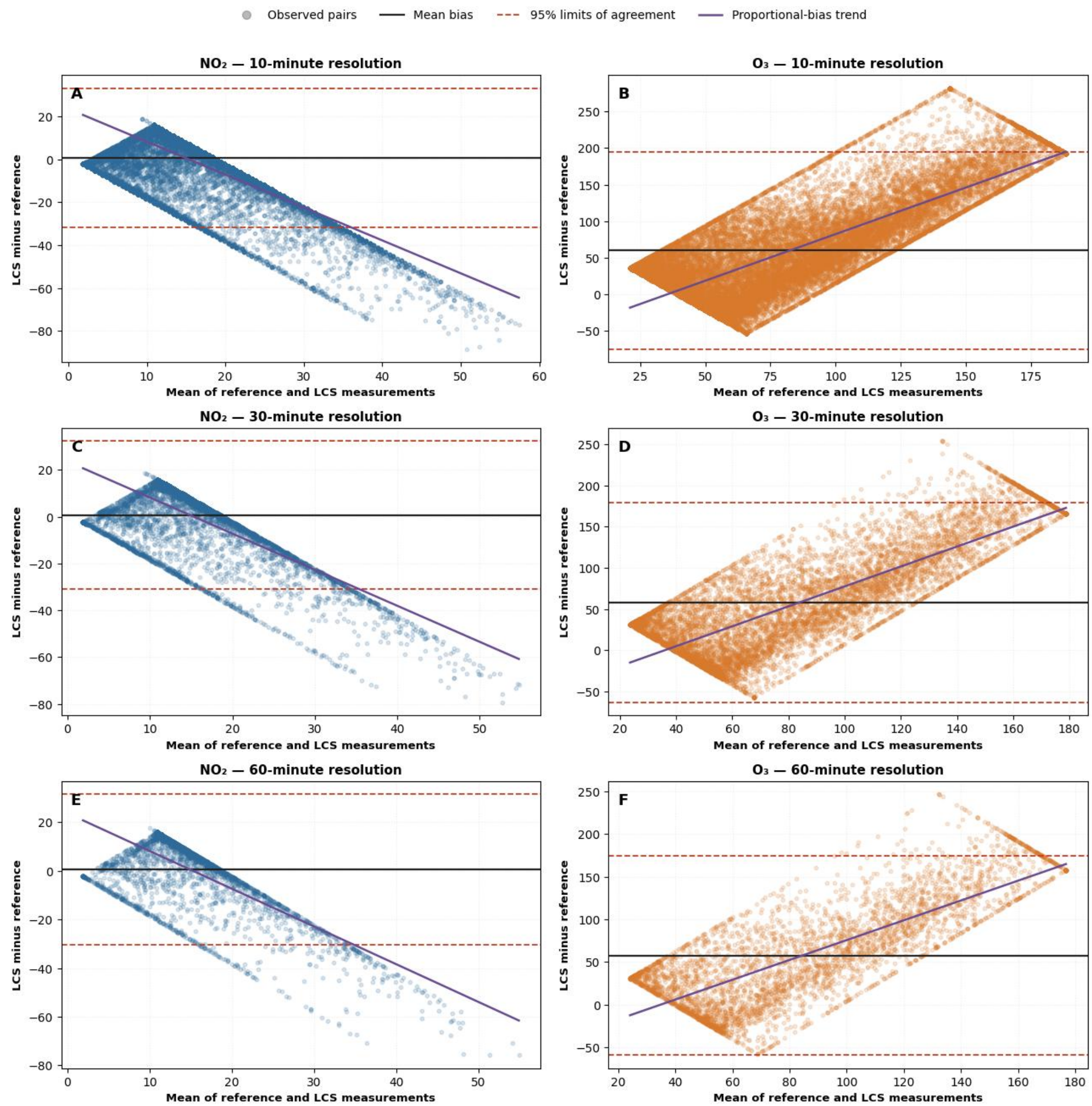

Figure 3. Bland–Altman Agreement: (A) NO_2, 10 min; (B) O_3, 10 min; (C) NO_2, 30 min; (D) O_3, 30 min; (E) NO_2, 60 min; (F) O_3, 60 min.

4. Discussion

The results demonstrate that the statistical behaviour of the low cost sensor measurement was significantly impacted by the temporal resolution, but the impact was different for NO_2 and O_3. NO_2 was weakly negatively correlated with all the reference measurements at all aggregation levels, and O_3 was moderately positively correlated with the reference measurements, with the correlation increasing as temporal aggregation increased from 10 to 60 minutes. The multivariate models helped to improve calibration, with a modest improvement with NO_2 and a more significant improvement with O_3. The best calibration for O_3 in the model was at 30 minutes, with a multivariate R^2 value of 0.382, and the error metrics were lower than the simple model. Temporal aggregation thus made for improved statistical stability but did not have a uniform effect on the benefits. The residual diagnostic also indicated that the ordinary regression assumptions were not adequately met as heteroscedasticity and high serial dependence were observed in all of the fitted models. Thus, it was crucial to obtain more reliable standard errors and confidence intervals through the use of HAC-adjusted inference.

The enhanced O_3 response in multivariate calibration is in line with the fact that pollutant behavior seems to be tightly coupled to meteorological variability and to larger-scale atmospheric processes. Climatic factors are found to explain significant percentage of variations in pollutants in different seasons and environmental conditions, which helps to incorporate temperature and humidity as covariate in the calibration models (Morshed et al., 2024). The present results show that by adjusting the environment, the calibration of the sensors can be enhanced even if the sensor-reference relationship is only moderate. The use of distributed low-cost networks has highlighted the need to consider influences from individual sensors and environment in addition to the raw output of the sensors (DeSouza et al., 2022). The aggregation over time also decreased a few error and variability metrics. The NO_2 calibration RMSE values decreased with time from 14.280 to 13.579 at 10 and 60 minutes, respectively, and the O_3 calibration RMSE values decreased the most between 10 and 30 minutes. This is common in many pollutant sensing applications where the concentration of pollutants is not constant over time, and where averaging over time can mask the effects of short-term variability in the distributed sensing environment (Coleman & Meggers, 2018). Temporal detail in sensing studies has also shown that information at a smaller scale can be gained by temporal detail, with some loss of statistical stability due to the temporal amalgamation of data (Weissert et al., 2019).

The agreement analysis is a dimension which goes beyond correlation. The mean bias for NO_2 was small, but the limits of agreement were wide, and the slope of the proportional bias was negative, while the mean bias for O_3 was high and positive and decreased with aggregation. The results show that the improvements in association do not necessarily lead to the exchangeability of low-cost and reference measurements. Careful reconciliation of systematic differences between measurement sources prior to fusion or use in an air-quality mapping framework is also essential for distributed sensor-based mapping of air quality (Schulte et al., 2020). Calibration studies of portable low-cost sensor stations have similarly shown that sensor performance can vary with pollutant type, environmental conditions, and deployment context (Tarazona Alvarado et al., 2024). Chronological validation gave a more conservative indication of calibration performance than did in-sample statistics. For all resolutions despite a more robust in-sample fit, the respective test R^2 was close to zero for NO_2 and negative for O_3. This separation suggests that there did not exist a perfect stability of the calibration relationships during the later observation period. Test RMSE and MAE were decreased by temporal aggregation, but it did not remove this instability. Studies with a sensor network have indicated that the spatial and temporal coverage can make a significant contribution to the spatial and temporal representation of concentration fields, but with this increased resolution comes problems of temporal transferability, changing environmental conditions, and sensor uniformity (Ahangar et al., 2019). The present results thus indicate that temporal aggregation can have a positive effect on the numerical stability of the model without being a guarantee of the model's temporal generalization.

These results have implications for practice and statistics. For the O_3, the 30-minute interval seemed to give the best compromise with a larger reduction in calibration error, multivariate R^2, and slightly larger test error and narrower agreement limits, than the 60-minute interval. The difference between the single and aggregated results was small for NO_2 suggesting more calibration structure is needed with this pollutant. Therefore, for monitoring applications, the temporal resolution should be chosen based on the statistics to be evaluated and not necessarily as a measure of better performance. Some limitations should be noted. The observation of the calibration patterns in this analysis was limited to one monitoring setting and two gaseous pollutants and may not apply directly to other sites or sensor configurations. To keep the model simple and interpretable, the calibration models did not include the most common nonlinear and seasonal effects. There is the potential for future work to consider nonlinear calibration functions, rolling models, adaptive models, and external validation for further monitoring sites with a focus on agreement and temporal robustness.

5. Conclusion

The impacts of the temporal resolution were very significant on the calibration behavior and agreement of the low-cost NO_2 and O_3 sensors, but were different depending on the pollutant. NO_2 showed weak negative correlations with reference measurements across all three resolutions, whereas O_3 showed stronger positive correlations that increased with temporal aggregation. The calibration model with a multivariate (MV) approach using temperature and relative humidity (RH) was always superior to the calibration model with sensor-only data, especially for O_3. The 30-minute resolution yielded the best in-sample calibration of O_3, while the 60-minute aggregation generally improved error reduction and decreased the limits of agreement. HAC-robust inference corroborated the results and proved the main predictor effects statistically significant even in the presence of heteroscedasticity and serial dependence. Consistent proportional bias was also observed across all pollutants using Bland–Altman analysis, suggesting that association and accuracy of measurements were not necessarily correlated. Chronological validation indicated a limited temporal generalization, especially for O_3, highlighting the importance of out-of-sample temporal validation of calibration models. Overall, the performance of the low-cost sensors is influenced by the pollutant type, environmental adjustment and temporal resolution, and calibration strategies should consider these factors for their use in operational air-quality monitoring.

References

1.Ahangar, F. E., Freedman, F. R., & Venkatram, A. (2019). Using low-cost air quality sensor networks to improve the spatial and temporal resolution of concentration maps. International Journal of Environmental Research and Public Health, 16(7), 1252.

2.Baklanov, A., & Zhang, Y. (2020). Advances in air quality modeling and forecasting. Global Transitions, 2, 261-270.

3.Bi, J., Carmona, N., Blanco, M. N., Gassett, A. J., Seto, E., Szpiro, A. A., ... & Sheppard, L. (2022). Publicly available low-cost sensor measurements for PM2. 5 exposure modeling: Guidance for monitor deployment and data selection. Environment international, 158, 106897.

4.Cavaliere, A., Brilli, L., Andreini, B. P., Carotenuto, F., Gioli, B., Giordano, T., ... & Gualtieri, G. (2023). Development of low-cost air quality stations for next-generation monitoring networks: calibration and validation of NO 2 and O 3 sensors. Atmospheric Measurement Techniques, 16(20), 4723-4740.

5.Clements, A. L., Griswold, W. G., Rs, A., Johnston, J. E., Herting, M. M., Thorson, J., ... & Hannigan, M. (2017). Low-cost air quality monitoring tools: from research to practice (a workshop summary). Sensors, 17(11), 2478.

6.Coleman, J. R., & Meggers, F. (2018). Sensing of indoor air quality—characterization of spatial and temporal pollutant evolution through distributed sensing. Frontiers in Built Environment, 4, 28.

7.DeSouza, P., Kahn, R., Stockman, T., Obermann, W., Crawford, B., Wang, A., ... & Kinney, P. (2022). Calibrating networks of low-cost air quality sensors. Atmospheric Measurement Techniques, 15(21), 6309-6328.

8.Doğan, N. Ö. (2018). Bland-Altman analysis: A paradigm to understand correlation and agreement. Turkish journal of emergency medicine, 18(4), 139-141.

9.Garcia, A., Saez, Y., Harris, I., Huang, X., & Collado, E. (2025). Advancements in air quality monitoring: a systematic review of IoT-based air quality monitoring and AI technologies. Artificial Intelligence Review, 58(9), 275.
10.Karagulian, F., Barbiere, M., Kotsev, A., Spinelle, L., Gerboles, M., Lagler, F., ... & Borowiak, A. (2019). Review of the performance of low-cost sensors for air quality monitoring. Atmosphere, 10(9), 506.
11.Liu, X., Jayaratne, R., Thai, P., Kuhn, T., Zing, I., Christensen, B., ... & Morawska, L. (2020). Low-cost sensors as an alternative for long-term air quality monitoring. Environmental research, 185, 109438.
12.Lukyanets, A., Gura, D., Savinova, O., Kondratenko, L., & Lushkov, R. (2023). Industrial emissions effect into atmospheric air quality: mathematical modeling. Reviews on Environmental Health, 38(2), 385-393.
13.Meneses Albalá, E., Guillem Montalbán Faet, Felici-Castell, S., Pérez Solano, J. J., & Segura-Garcia, J. (2025). Air Quality Data from Regulatory AQMS and Low-Cost Sensors in Bulevar Sur, Valencia (July 08 - November 18, 2023) (Version 1.0.0) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.17018565
14.Morawska, L., Thai, P. K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., ... & Williams, R. (2018). Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone?. Environment international, 116, 286-299.
15.Morshed, S. R., Fattah, M. A., Kafy, A. A., Alsulamy, S., Almulhim, A. I., Shohan, A. A. A., & Khedher, K. M. (2024). Decoding seasonal variability of air pollutants with climate factors: A geostatistical approach using multimodal regression models for informed climate change mitigation. Environmental Pollution, 345, 123463.
16.Okorn, K., & Iraci, L. T. (2024). An overview of outdoor low-cost gas-phase air quality sensor deployments: current efforts, trends, and limitations. Atmospheric Measurement Techniques, 17(21), 6425-6457.
17.Rybarczyk, Y., & Zalakeviciute, R. (2018). Machine learning approaches for outdoor air quality modelling: A systematic review. Applied Sciences, 8(12), 2570.
18.Schulte, N., Li, X., Ghosh, J. K., Fine, P. M., & Epstein, S. A. (2020). Responsive high-resolution air quality index mapping using model, regulatory monitor, and sensor data in real-time. Environmental Research Letters, 15(10), 1040a7.
19.Tarazona Alvarado, M., Salamanca-Coy, J. L., Forero-Gutièrrez, K., Núñez, L. A., Pisco-Guabave, J., Escobar-Diaz, F., & Sierra-Porta, D. (2024). Assessing and monitoring air quality in cities and urban areas with a portable, modular and low-cost sensor station: calibration challenges. International Journal of Remote Sensing, 45(17), 5713-5736.
20.Tella, A., Balogun, A. L., & Faye, I. (2021). Spatio-temporal modelling of the influence of climatic variables and seasonal variation on PM10 in Malaysia using multivariate regression (MVR) and GIS. Geomatics, Natural Hazards and Risk, 12(1), 443-468.
21.Weissert, L. F., Alberti, K., Miskell, G., Pattinson, W., Salmond, J. A., Henshaw, G., & Williams, D. E. (2019). Low-cost sensors and microscale land use regression: Data fusion to resolve air quality variations with high spatial and temporal resolution. Atmospheric environment, 213, 285-295.